

Práctica 1

Tipología y ciclo de vida de los datos

Sebastian Maya

Evolución de precios de la electricidad en Europa

Descripción

Este dataset comprende el historial del precio de la electricidad para cada semestre desde el año 2007 hasta el 2021, indicando el coste en euros por kilovatio-hora registrado para los consumidores domésticos de los países que forman la Unión Europea.

Contexto

La coyuntura política, económica y social actual ha llevado a un aumento reciente en los precios de la electricidad y otros bienes energéticos básicos. Es por eso que considero relevante hacer un estudio de los precios de electricidad desde 2007 hasta la actualidad. He elegido empezar el dataset desde 2007 porque el cálculo de consumo y su referencia se basa en bandas de consumo predefinidas y valores promedio para cada semestre (anteriormente se asignaba un precio fijo estático para cada semestre[1]).

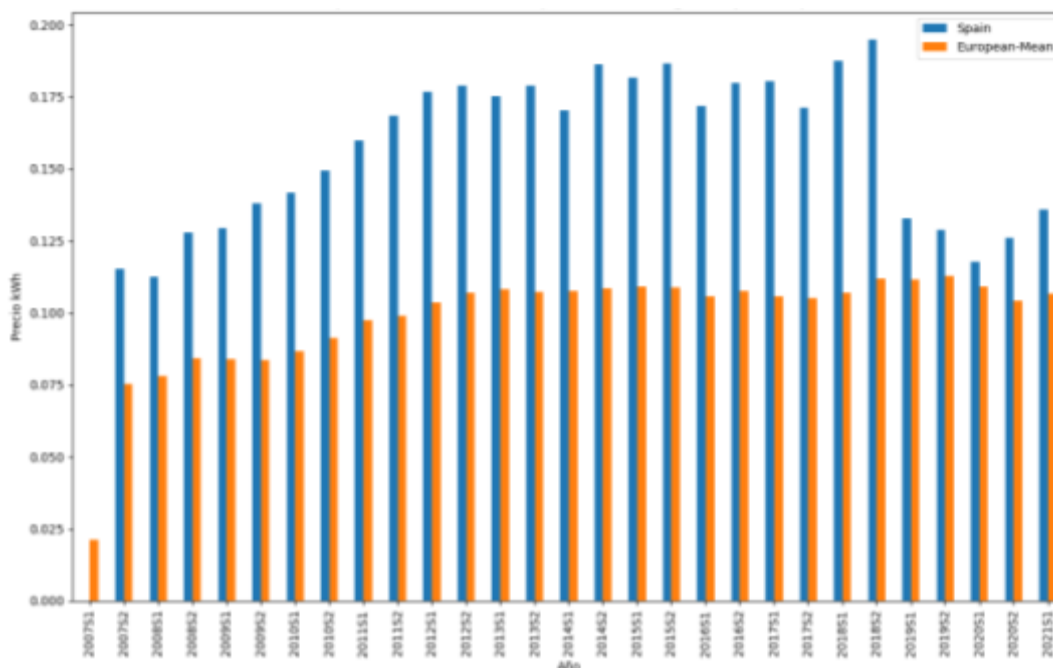
Estos datos dependen de muchos factores que varían en diferentes regiones. Para una representación accesible de los datos, los presenté a nivel nacional; los países elegidos son los estados miembros de la UE, varios países candidatos/potenciales y los países de la AELC Noruega y Liechtenstein.

La web elegida es Eurostat, la oficina europea de estadística[2]. Al tratarse de un organismo de la unión europea podemos asegurar unos datos fiables y la estandarización de su proceso de colección.

Para contextualizar los datos es necesario saber que los usuarios domésticos son aquellos con consumo anual medio entre 2.500 kWh y 5000 kWh. Así como que los precios publicados son sin impuestos ni gravámenes.

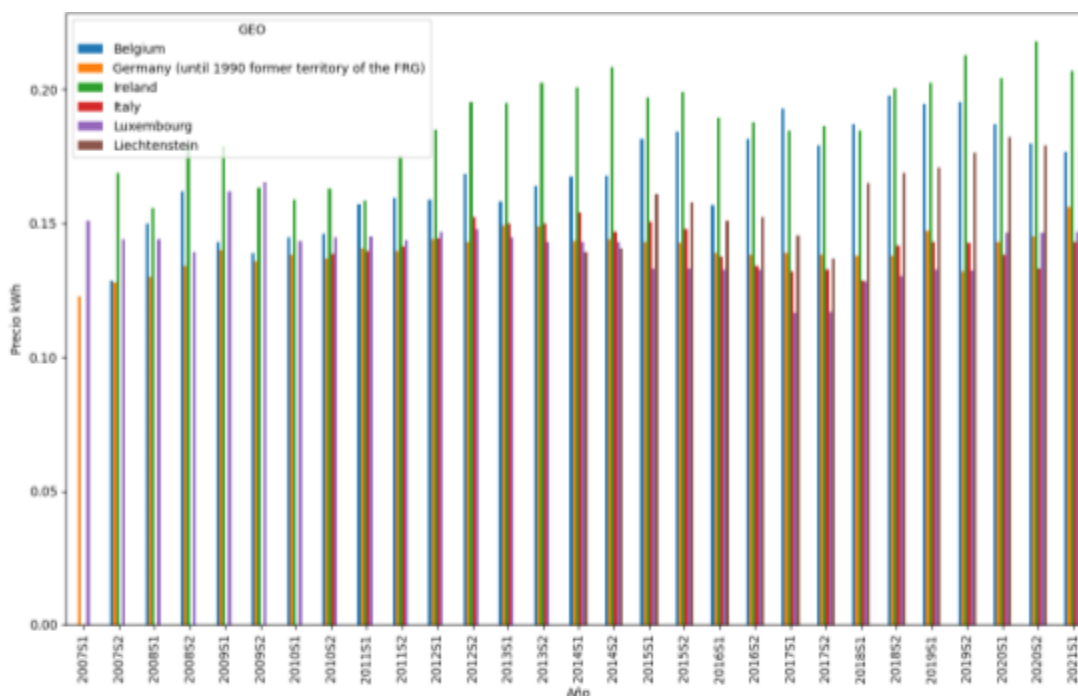
Representación gráfica

El siguiente diagrama es una buena representación del objetivo de este dataset, nos expone una comparativa histórica de precios entre la tendencia española y europea. Este gráfico sirve como referencia ya que podemos hacer un análisis similar para cualquier otro país o conjunto de países de la EU.



Comparativa de la evolución de precios de la energía Europa vs. España

En el siguiente gráfico podemos ver un análisis más detallado con otra perspectiva; la variedad de países en el dataset nos permite elaborar diferentes tipos de ranking. Un ejemplo es el top 5 que vemos a continuación.



Comparativa de la evolución de precios de la energía en Europa para los países más caros

Contenido

Para cada país de la unión europea, se reflejan los datos por año y semestre:

- Año inicial 2007 y año final 2021.
- Tenemos el semestre 1 y 2 (marcados como s1 y s2) para cada año. Los precios del semestre 1 son precios promedio entre el 1 de enero y el 30 de junio de cada año, los precios del semestre 2 son precios promedio entre el 1 de julio y el 31 de diciembre de cada año.

El listado de países incluidos es el siguiente:

- Bélgica
- Bulgaria
- Chequia
- Dinamarca
- Alemania (hasta 1990 antiguo territorio de la RFA)
- Estonia
- Irlanda
- Grecia
- España
- Francia
- Croacia
- Italia
- Chipre
- Letonia
- Lituania
- Luxemburgo
- Hungría
- Malta
- Países Bajos
- Austria
- Polonia
- Portugal
- Rumania
- Eslovenia
- Eslovaquia
- Finlandia
- Suecia
- Islandia
- Liechtenstein
- Noruega
- Reino Unido

- Montenegro
- Macedonia del Norte
- Albania
- Serbia
- Turquía
- Bosnia y Herzegovina
- Kosovo (según la Resolución 1244/99 del Consejo de Seguridad de las Naciones Unidas)
- Moldavia
- Ucrania
- Georgia
- Unión Europea - 27 países desde 2020
- Zona euro (EA11-1999, EA12-2001, EA13-2007, EA15-2008, EA16-2009, EA17-2011, EA18-2014, EA19-2015)

Para recopilar los datos, Eurostat proporciona modelos de cuestionarios a todos los proveedores de información. Estas fuentes son institutos nacionales de estadística, ministerios, agencias de energía o en caso de monopolios empresas eléctricas individuales que compilan precios reales facturados que pagan los usuarios finales. Eurostat se encarga de calcular los agregados de la UE ponderando los precios nacionales con el último consumo disponible (se puede encontrar más información en [3]) para el sector doméstico.

Agradecimientos

Los datos han sido recolectados desde el portal de la oficina estadística de la comisión europea, Eurostat, concretamente de la base de datos “*nrg_pc_204*”. Esta base de datos es ampliamente utilizada por diferentes organismos para presentar estudios sobre los precios entre países, ejemplos de ello son los presentados en [4] como “Cómo difiere el peso de los impuestos y gravámenes entre los Estados miembros”, “La mayor caída de los precios de la electricidad en los Países bajos, Chipre y Lituania” o el “Precio de la electricidad en estándar de poder adquisitivo (PPS)”, estos y otros análisis se publican de manera recurrente.

Se ha creado un programa en Python que utiliza librerías como Selenium, request, Pandas, matplotlib, entre otros; y técnicas avanzadas de web scraping para extraer la información. Se ha hecho especial énfasis en la librería Selenium [5] que ha permitido cargar la información más compleja generada a partir de JavaScript, jQuery, mensajes asíncronos y ventanas emergentes, siendo esto el punto más difícil de la extracción de información.

Se han seguido los principios éticos y legales del tratamiento de datos. El Eurostat sigue el reglamento (CE) no 223/2009 sobre estadísticas europeas de 11 de marzo de 2009. Así podemos saber que no divulga datos que puedan estar sujetos a confidencialidad comercial. Los datos confidenciales solo son accesibles por Eurostat, sin embargo, los datos de precios se pueden utilizar con fines estadísticos y no impedirá su publicación de forma agregada (lo que no permite identificar transacciones individuales).

Por otra parte las políticas de acceso de usuarios destacan lo siguiente:

"Política de acceso de usuarios: De conformidad con el marco jurídico comunitario y el Código de buenas prácticas de las estadísticas europeas, Eurostat difunde estadísticas europeas en el sitio web de Eurostat respetando la independencia profesional y de una manera objetiva, profesional y transparente en la que se trata a todos los usuarios"[1]

Para finalizar, también sigo los principios legales en cuanto al uso de los datos ya que Eurostat tiene una licencia de creative commons 4.0. que veremos más adelante en el apartado de licencia.

Inspiración

En los últimos años se percibe un incremento constante en los precios de los suministros básicos, especialmente el de la electricidad. Son muchos los artículos publicados[6] para alertar sobre este incremento constante y hablar de récords históricos, o para hacer campaña en contra del partido político que gobierna, abriendo un debate interminable. Pero ¿es realmente esta sensación una consecuencia proveniente del incremento de los precios de los mayoristas? ¿Son los impuestos cada vez más asfixiantes? ¿o simplemente son un resultado de malas políticas? Mi objetivo es proporcionar datos imparciales con los que la ciudadanía pueda sacar conclusiones reales sobre el precio en un mercado tan regulado como el europeo y descubrir las respuestas a estas y otras preguntas relacionadas con los precios de la electricidad.

El objetivo de este trabajo, y como consiguiente la generación de este dataset, se alinea con los objetivos antes descritos por Eurostat; presentar datos de manera objetiva, profesional y transparente para impulsar la oferta del mercado, reducir los precios para los consumidores y fomentar un aumento de la competencia, la inversión y la productividad que conduzca a un aumento de la eficiencia económica.

Licencia

La licencia más adecuada para este dataset es la “CC BY-SA 4.0 License”, principalmente porque los contenidos publicado en el sitio web de Eurostat se publican bajo la licencia “CC BY 4.0” (a menos de que se indique lo contrario como con avisos específicos de derechos de autor[7]) y además esta licencia es la más conveniente para este dataset por lo siguiente:

1. **Reglas para compartir:** *Se permite copiar y distribuir el material en cualquier medio o formato.* Esto ayuda a una distribución más sencilla entre posibles usuarios interesados en los datos.
2. **Reglas para adaptar:** *Se permite remezclar, transformar y crear a partir del material para cualquier finalidad incluida la comercial.* Esto permite que más usuarios, profesionales o no, tengan acceso a los datos y puedan utilizarlos de manera libre en su trabajo diario.
3. **Términos de atribución o reconocimiento:** *Se debe dar reconocimiento o crédito a la fuente e indicar si se han realizado cambios, pero no de forma que sugiera que su uso tiene el apoyo del licenciante.* Así pues, en este proyecto se reconoce el trabajo de la Unión Europea en la colección y presentación de los datos.
4. **Otras medidas:** *No se pueden aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia[8].* Esto indica que las contribuciones posteriores deben publicarse bajo la misma licencia, respetando el trabajo del autor original.

Código

En [9] se encuentra el código fuente para la extracción de los datos, una explicación de las clases que lo conforman, requerimientos y cómo ejecutarlo; y finalmente el archivo con el dataset generado.

Dataset in Zenodo

En el siguiente enlace se encuentra la publicación del dataset obtenido en Zenodo

- <https://doi.org/10.5281/zenodo.5652062>

Referencias

1. https://ec.europa.eu/eurostat/cache/metadata/en/nrg_pc_204_esms.htm
2. <https://es.wikipedia.org/wiki/Eurostat>
3. <https://ec.europa.eu/eurostat/web/energy/methodology/prices>
4. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_price_statistics#Electricity_prices_for_household_consumers
5. <https://www.selenium.dev/selenium/docs/api/py/api.html>
6. https://www.elconfidencial.com/economia/2021-09-03/luz-gas-gasolina-suministros-precio-dia-subida_3262058/
7. https://ec.europa.eu/info/legal-notice_es
8. https://creativecommons.org/licenses/by-sa/4.0/deed.es_ES
9. https://github.com/damaro05/Practica1_Tipologia_y_ciclo_de_vida_de_los_datos

Contribuciones

Contribuciones	Firmas
Investigación previa	Sebastian Maya
Redacción de las respuestas	Sebastian Maya
Desarrollo del código	Sebastian Maya